S ummary

Project Topic: Voting Behavior

The main idea of this project is to analyze LLM's ability in predicting voting trends based on voter characteristics. We provide the LLM with information about voters including demographics, previous voting behavior, and any known political affiliations.

The goal of this experiment is to test whether the model responds sensibly to demographic cues, uncertainty, and ambiguous signals.

Below is a summary of the first round of experimenting and individual summaries of each response:

Overall Summary

The prompts used were varied in the kinds of voter information they were provided and the kinds of outputs requested. The responses of the first few prompts were straight forward and as expected. The responses became especially long for the last few prompts, especially when the model was asked a more open ended question (as expected).

To fine tune the inputs, one significant change is adjusting the inputs based on the kind of real world data we have available. We will specify and create more inputs based on the data, and also use the data to assess the accuracy and reliability of the outputs. Some examples of the types of data we have available are:

- https://www.sos.wa.gov/elections/data-research/election-data-and-maps/reports-data-and-statistics/voter-demographics

- https://www.sos.wa.gov/elections/data-research/reports-data-and-statistics/voter-participation-statistics

- https://www.census.gov/data/tables/time-series/demo/voting-and-registration/p20-587.html

One limitation with the real data we currently have for the U.S. is that while we have detailed county, city, and state data we are limited in the variety of demographics available. However we have also found some research indicating variables such as age, gender, and economic factors are most significant in influencing voting behaviors and these are the kinds of variables we have the best access too. Moving forward we will focus on these variables. We are also still looking for international data to use.

We have not decided on a plan to automate largescale data collection as we have been focusing on getting the model working and understanding how to use the LLM.

Our next steps will also focus on planning analysis methods, determining what types of outputs will best suit those methods, and adjusting inputs to provide those. The outputs we have received thus far are interesting in an exploratory sense but need to be refined for proper analysis.

Prompt 1: Basic Demographic Prediction

This prompt was meant to be a straightforward baseline. The voter profile includes clear demographic and geographic signals, along with consistent past voting behavior. The goal was to see if the model could recognize an "easy case" and give a reasonable prediction without overthinking it.

Summary of the response: The model predicted that the voter would likely vote Democratic. The explanation focused heavily on past voting behavior, which is usually the strongest indicator of future votes. It also pointed to Seattle's strong Democratic leaning and the voter's age, education level, and gender, all of which tend to correlate with Democratic support. The model acknowledged that party registration as Independent doesn't necessarily mean political neutrality, especially when past behavior is consistent. Overall, the response aligned with common political trends, which makes sense given how clear the signals were.

Prompt 2: Comparative Prediction

Instead of predicting how someone would vote, this prompt asked which voter is easier to predict. The goal was to see whether the model understands uncertainty and recognizes that some voters are more predictable than others.

Summary of the response: The model correctly identified Voter A as easier to predict. It explained that having a long, consistent voting history—especially at an older age—makes future behavior more predictable. In contrast, Voter B is a first-time voter with no history, making their behavior much harder to guess. The response explained why missing data matters and avoided claiming certainty for either voter. This shows the model can reason about predictability itself, not just outcomes.

Prompt 3: International Comparison

This prompt tested whether the model could adapt its reasoning to a non-U.S. political system and handle more complex issue-based tradeoffs. It also aimed to see how the model weighs conflicting concerns like economic pressure and immigration.

Summary of the response: The response gave a detailed analysis and ultimately suggested that the voter is more likely to move away from Labour and toward a Conservative or populist party. The model emphasized immigration as a key "deal-breaker" issue, even though the voter's education level and past voting behavior would normally suggest continued Labour support. It also discussed how cost-of-living concerns could push the voter in different directions depending on which party's solutions seem more credible.

Prompt 4: Minimal Voter Information

This prompt intentionally provided very little information and told the model not to make a prediction. The goal was to see whether the model could explain uncertainty and identify what information is missing instead of guessing.

Summary of the response: The model clearly stated that uncertainty was high and avoided making a prediction. It listed several types of missing data that would be important, such as party affiliation, key political issues, income, occupation, and past voting behavior. The response also explained why the existing information (age, location, education) was too broad to be predictive on its own.

Prompt 5: Explicit Uncertainty

This prompt directly asked the model to make a rough prediction while also explaining uncertainty and limitations. It was designed to see how well the model can reflect on its own weaknesses and the limits of demographic-based predictions.

Summary of the response: The model gave a very cautious prediction that the voter would slightly lean Democratic, mainly due to New York's overall political trend. However, it clearly labeled this prediction as extremely low confidence. The response included a long list of uncertainty sources, such as missing past voting behavior, lack of precise location, and unknown personal values. It also explained broader limitations, like the risk of using group-level trends to predict individuals.

Prompt 6: Conflicting Past Voting Behavior

This prompt focused on a voter with mixed past behavior (the voter voted for Republican), which makes them a classic swing voter. The goal was to see how the model handles conflicting signals and whether it adjusts its confidence appropriately.

Summary of the response: The model predicted a slight Republican lean but emphasized that confidence was low. It explained that the voter's party switch in 2020 suggests flexibility rather than loyalty, possibly driven by specific candidates rather than ideology. The response highlighted the importance of rural Arizona's political context and recent shifts among Hispanic voters in the region. At the same time, it stressed that this voter is highly persuadable and could easily swing again depending on the candidates and issues.